

Factor relationships and distributions

EXPLORATORY DATA ANALYSIS IN PYTHON



Izzy Weber

Curriculum Manager, DataCamp



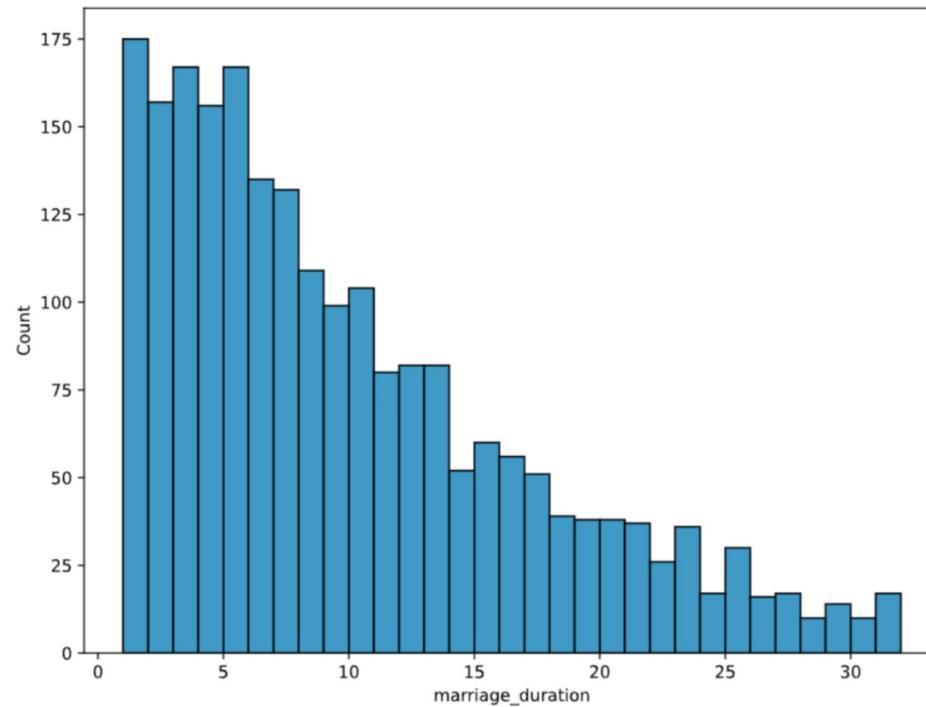
Level of education: male partner

```
divorce["education_man"].value_counts()
```

```
Professional      1313
Preparatory       501
Secondary         288
Primary           100
None              4
Other              3
Name: education_man, dtype: int64
```

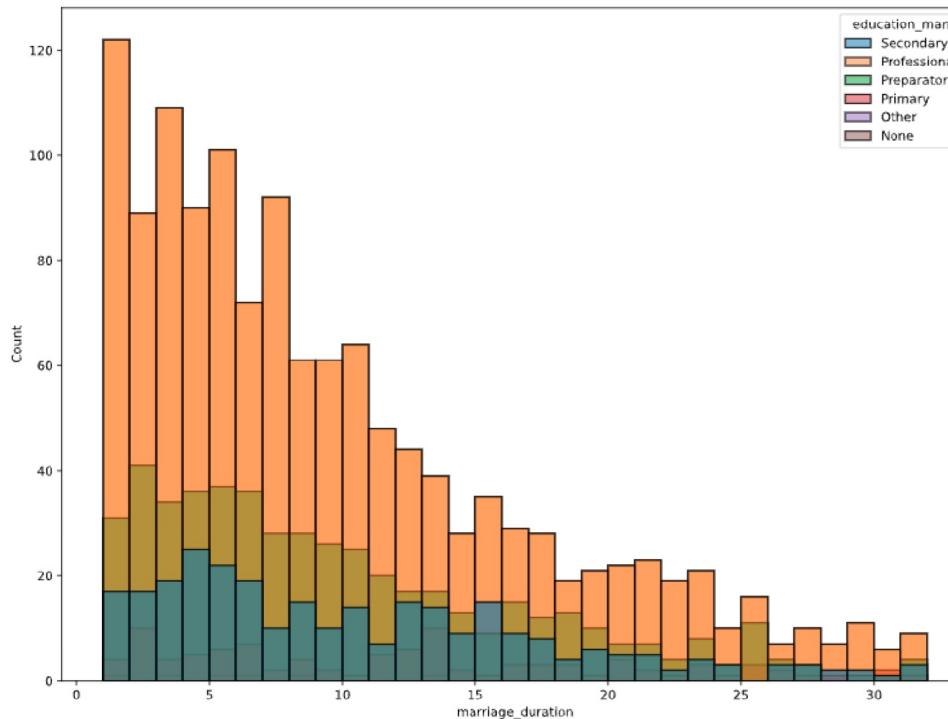
Exploring categorical relationships

```
sns.histplot(data=divorce, x="marriage_duration", binwidth=1)  
plt.show()
```



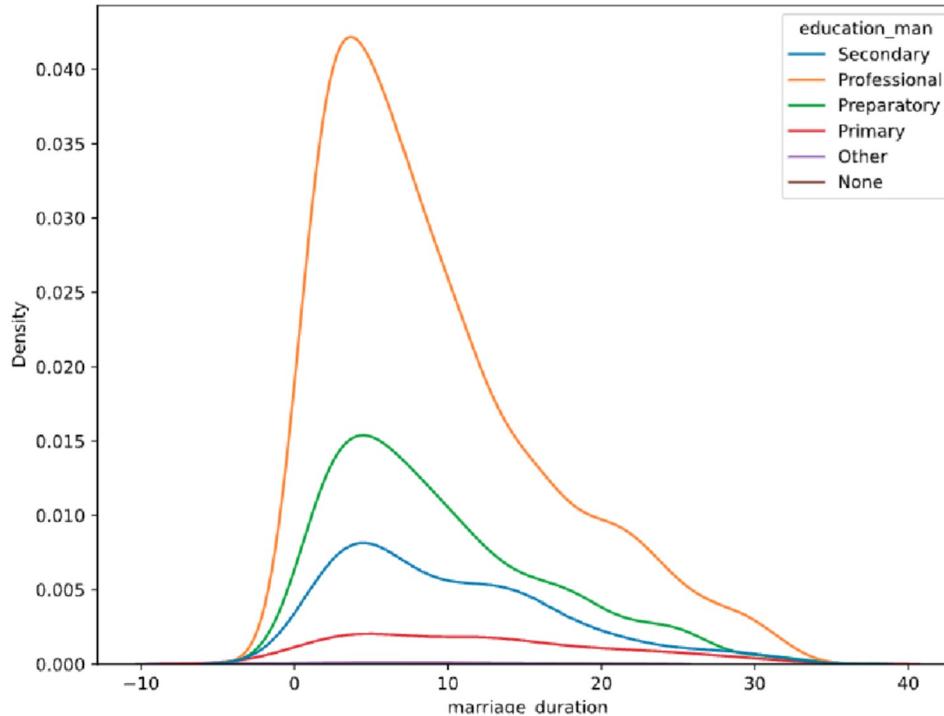
Exploring categorical relationships

```
sns.histplot(data=divorce, x="marriage_duration", hue="education_man", binwidth=1)  
plt.show()
```

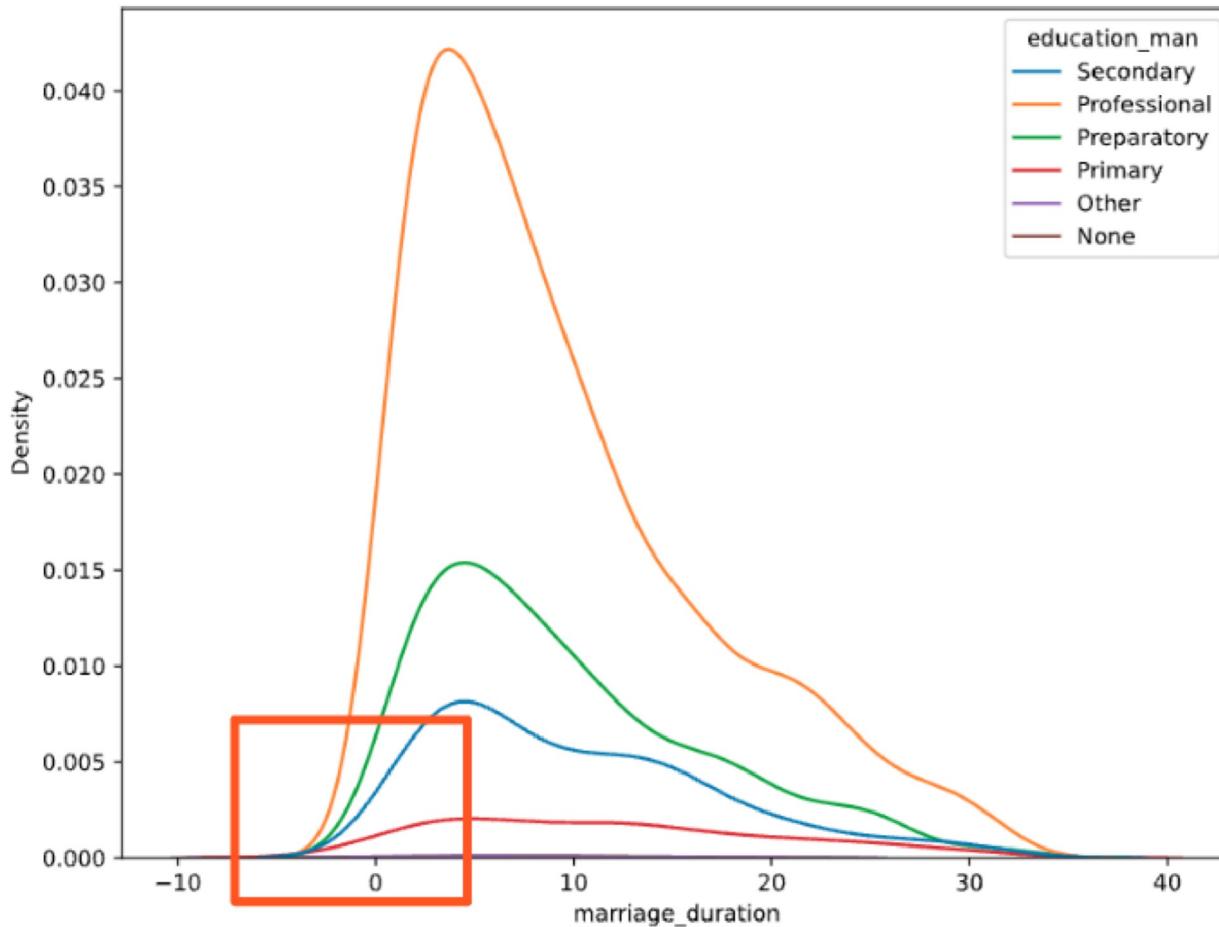


Kernel Density Estimate (KDE) plots

```
sns.kdeplot(data=divorce, x="marriage_duration", hue="education_man")  
plt.show()
```

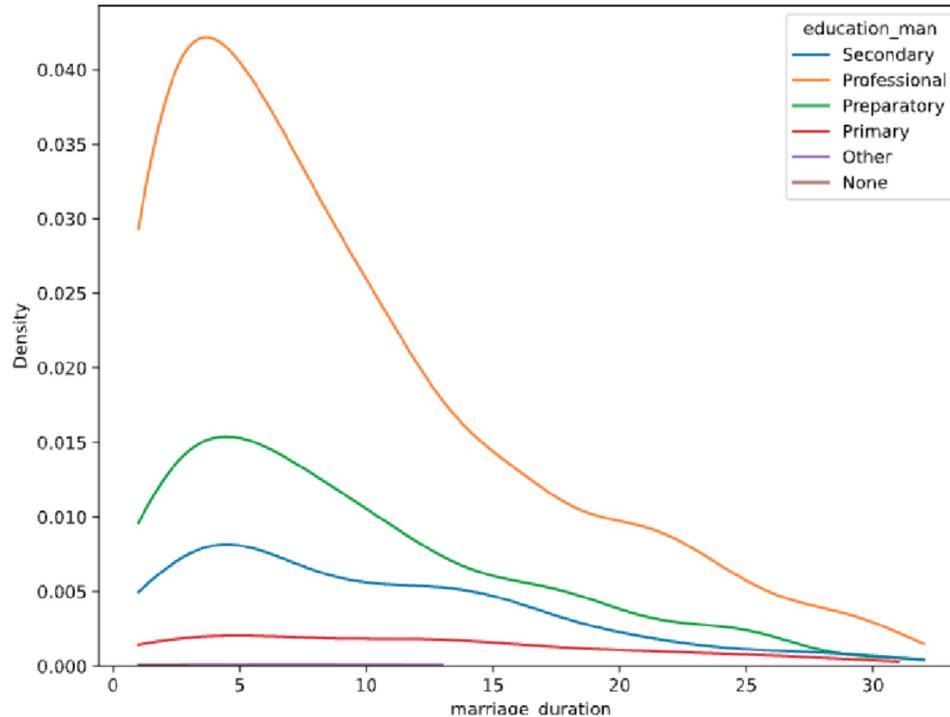


Kernel Density Estimate (KDE) plots



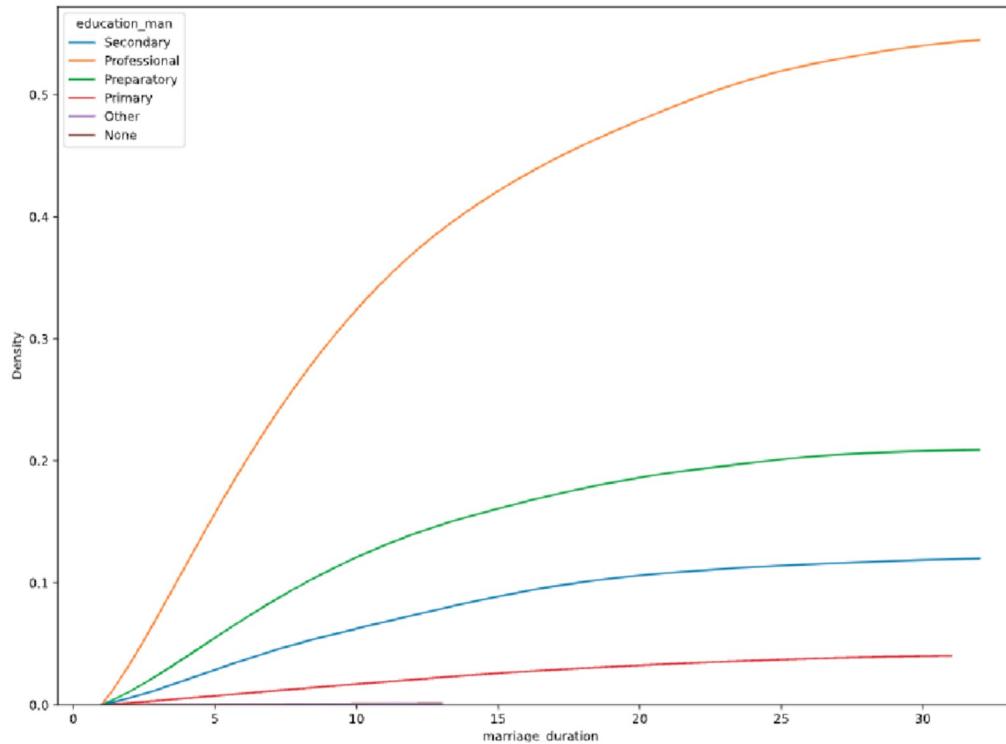
Kernel Density Estimate (KDE) plots

```
sns.kdeplot(data=divorce, x="marriage_duration", hue="education_man", cut=0)  
plt.show()
```



Cumulative KDE plots

```
sns.kdeplot(data=divorce, x="marriage_duration", hue="education_man", cut=0, cumulative=True)  
plt.show()
```



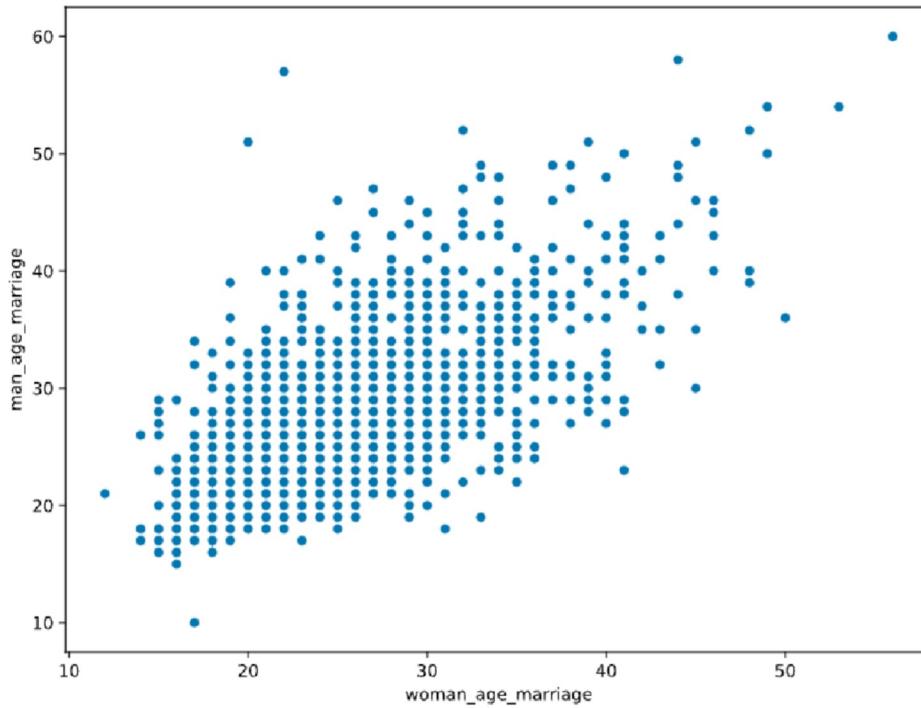
Relationship between marriage age and education

- Is there a relationship between age at marriage and education level?

```
divorce["man_age_marriage"] = divorce["marriage_year"] - divorce["dob_man"].dt.year  
divorce["woman_age_marriage"] = divorce["marriage_year"] - divorce["dob_woman"].dt.year
```

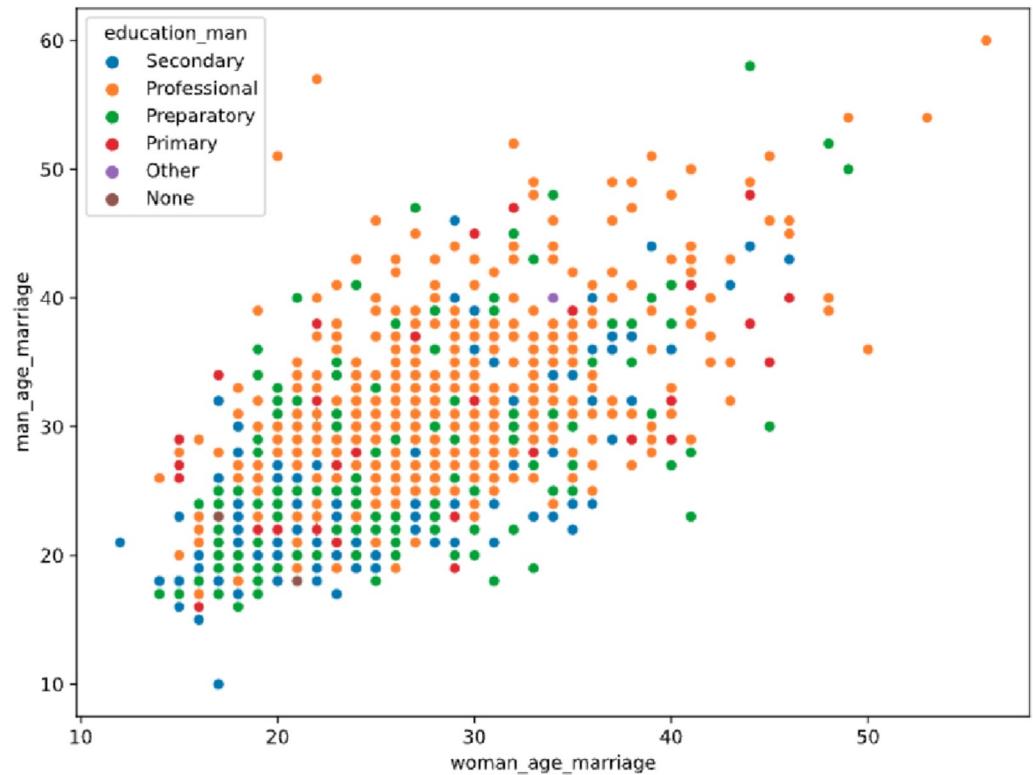
Scatter plot with categorical variables

```
sns.scatterplot(data=divorce, x="woman_age_marriage", y="man_age_marriage")  
plt.show()
```



Scatter plot with categorical variables

```
sns.scatterplot(data=divorce,  
                  x="woman_age_marriage",  
                  y="man_age_marriage",  
                  hue="education_man")  
  
plt.show()
```



Let's practice!

EXPLORATORY DATA ANALYSIS IN PYTHON

